Temporal Occupancy Grids: a Method for Classifying the Spatio-Temporal Properties of the Environment

Daniel Arbuckle Andrew Howard Maja Matarić

Robotics Research Laboratory, Computer Science Department, University of Southern California
arbuckle@usc.edu, ahoward@usc.edu, mataric@usc.edu

Abstract

This paper introduces the concept of a temporal occupancy grid *as a method for modeling and classifying spatial areas according to the time properties of their occupancy. The method extends the idea of occupancy grids[1] by considering occupancy over a number of different timescales. This paper presents the basic formalism and its implementation using planar laser rangefinders. It includes the results of a number of validation experiments, and an experiment in which we demonstrate the ability to locate doors in a real-world setting.*

1 Introduction

Robots and various automated systems have a strong need to know about their environment. The first step to this is mapping the static features of an area, but there is far more to know about an area than what shape it has. This paper is concerned with learning the motion patterns situated in and associated with an area.

Locations in an area can be classified by how likely they are to be occupied, as in the classic occupancy grid (OG) [1]. Temporal occupancy grids (TOGs) are an extension of this idea through the time dimension, allowing the classification of grid cells based on the time properties of their occupancy. Specifically, motions within the area under consideration have the property that an OG which represents a length of time less than or equal to the length that the motion remains in a specific grid cell will show a higher probability of that cell being occupied than will an OG spanning a longer timescale.

A temporal occupancy grid is essentially a matrix with two spatial dimensions, one time dimension and a number of additional dimensions equal to the number of different timescales being considered. Thus, each layer of the TOG is essentially several OGs, each representing a different period of time leading up to the specific moment which the layer describes.

TOGs can differentiate between different patterns of occupancy, even when the absolute probability of occupancy is the same. At any given moment, it is simple to classify the occupancy of a grid cell as "occupied on all timescales" (i.e., background), "occupied on no timescales," or as occupied on some combination of timescales under consideration.

Objects which move between cells at different rates leave distinctly different traces in a temporal occupancy grid. This makes it quite simple to separate the background from the more temporary occupancies. Traffic patterns, including which areas are used by swiftly moving objects and which by slower ones, can be extracted from a TOG. Mobile robots can apply these data to localization and navigation. TOGs representing the interactions of several people with each other and with the space can potentially be used to classify the type of interaction that occurred. It is hoped that abnormal uses of an area will be captured effectively by a TOG.

In the rest of the paper we present the theoretical background for TOGs and describe how they were implemented and applied. We validate the method by applying it to automated analysis of scripted human activity. The occupancy data are gathered with planar laser scanners, and analyzed at multiple timescales. The results provide dynamic and static analysis of the area through the use of TOGs.

2 Related Work

Our approach draws heavily on the body of work concerning classical OGs, particularly Elfes[1]. Occupancy grids provide the basic foundation upon which this work is built. Works in multi-object tracking also bear some relevance to TOGs. Sato and Aggarwal[2] approached the problem of classifying agent interactionss, using multiple cameras and computer vision techniques. Mittal and Davis[3] worked with the unification of input from a wide baseline array of cameras, and used this unification for person tracking. These papers provide some background to the problem of unifying several sensors in the context of human activity modeling. Gern & Gilles[9] and Olson[10], among others, used lasers to build OGs. Their work supports the use of laser data for generating occupancy grids. Schulz, Burgard, Fox & Cremers[5], Chang

& Gong[6], and Mahler[8] are among those who have dealt with modeling of human activity embedded in an area. The method we present in this paper is a new alternative to addressing this problem.

Isolating unusual events in an area has been studied as a computer vision problem, as in Collins, Lipton, Kanade, Fujiyoshi, Duggins, Tsin, Tolliver, Enomoto, Hasegawa, Burt and Wixson[11] and Flinchbaugh & Bannon[12]. The method we propose addresses the problem in a different and hopefully synergistic way.

3 Method

3.1 Occupancy Grids

Occupancy grids divide the world into a set of cells, and attempt to determine, for each cell, the probability that the cell is either *occupied* or *empty*. This determination is made by applying Bayes Law to a series of observations.

Let $s_{i,j} \in \{occ, emp\}$ denote the state of the cell at (i,j), let m_t denote an observation made at time t, and let $(m_t, m_{t-1}..., m_1)$ denote a series of such observations. Our aim is to determine the probability $p(s_{ij} \mid m_t, ..., m_1)$, i.e., the probability that cell (i,j) is in state $s_{i,j}$, given the series of observations $(m_t, ..., m_1)$. Under the assumption that observations are statistically independent, we can determine this probability using the incremental form of Bayes Rule:

$$p(s_{i,j} \mid m_t, ..., m_1) = \frac{p(m_t \mid s_{i,j}) p(s_{i,j} \mid m_{t-1}, ..., m_0)}{p(m_t)} \tag{1}$$

where $p(m_t \mid s_{i,j})$ is the probability of obtaining the observation m_t, given that the cell (i,j) is in state $s_{i,j}$, and $p(s_{i,j} \mid m_{t-1}, ..., m_0)$ is the probability that the cell is in state $s_{i,j}$, given the series of observations $(m_{t-1}, ..., m_1)$. The first of these terms is often referred to as the *sensor model*, while the latter is an incremental *prior probability*. The denominator in the above equation measures the probability of obtaining the measurement m_t, and is effectively a normalization constant.

Since $s_{i,j}$ can take only two values, we can conveniently re-write Equation 1 in terms of *likelihoods* (probability ratios):

$$\frac{p(s_{i,j} = occ \mid m_t, ..., m_0)}{p(s_{i,j} = emp \mid m_t, ..., m_0)} = \frac{p(m_t \mid s_{i,j} = occ)}{p(m_t \mid s_{i,j} = emp)} \frac{p(s_{i,j} = occ \mid m_{t-1}, ..., m_1)}{p(s_{i,j} = emp \mid m_{t-1}, ..., m_1)} \tag{2}$$

This expression can be further simplified by defining a pair of *log-likelihoods*, as follows. Let $O_{i,j,t}$ denote the *occupancy value* for cell (i,j) at time t:

$$O_{i,j,t} = \log \frac{p(s_{i,j} = occ \mid m_t, ..., m_1)}{p(s_{i,j} = emp \mid m_t, ..., m_1)} \tag{3}$$

and let $R_{i,j,t}$ denote the *log-likelihood sensor model* for cell (i,j) at time t

$$R_{i,j,t} = \log \frac{p(m_t \mid s_{i,j} = occ)}{p(m_t \mid s_{i,j} = emp)} \tag{4}$$

Substituting these definitions into Equation 2, we obtain:

$$\begin{aligned} O_{i,j,t} &= R_{i,j,t} + O_{i,j,t-1} \\ &= \sum_{0 < t' \leq t} R_{i,j,t'} \end{aligned} \tag{5}$$

Note that we assume equal a priori probabilities and hence the initial occupancy value for all cells is zero.

The sensors used in this work are SICK LMS-200 scanning laser range-finders. These devices are capable of returning range readings out to a distance of 8 m over a 180 degree swath. These range measurements are accurate to about 1 cm, and the angular resolution is approximately 1 degree.

Sensor models for this device are relatively easy to construct. Each observation m_t consists of a set of range-and-bearing measurements recorded by one or more lasers; if we assume that these measurements are statistically independent, we can update the grid separately for each range-and-bearing pair.

Let (r, ϕ) denote the range and bearing recorded by some laser k at time t, and let $(\Delta r, \Delta\phi)$ be the uncertainty associated with this measurement. We use the following simplified sensor model:

$$R_{i,j,t} = \begin{cases} -1 & \text{if } r_{i,j,k} < r - \Delta r \text{ and } |\phi_{i,j,k} - \phi| < \Delta\phi \\ +1 & \text{if } |r_{i,j,k} - r| < \Delta r \text{ and } |\phi_{i,j,k} - \phi| < \Delta\phi \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $r_{i,j,k}$ and $\phi_{i,j,k}$ denote the range and bearing of cell (i,j) relative to laser k. This sensor model captures three fairly intuitive cases. First, there are cells that lie along the measured bearing ϕ, but are closer to the laser than the measured range r. For these cells, the probability of obtaining the measurement (r, ϕ) is much higher if the cell is empty than if it is occupied; hence $R_{i,j,t} < 0$. Second, there are cells that lie along the measured bearing ϕ and are at the measured range r. For these cells, the probability of obtaining the measurement (r, ϕ) is much higher if the cell is occupied than if it is empty; hence $R_{i,j,t} > 0$. Finally, there are cells that do not lie along the measured bearing, or are further from the laser than the measured range. For these cells, the probability of obtaining the measurement (r, ϕ) is the same, irrespective of whether the cell is occupied; hence $R_{i,j,t} = 0$.

3.2 Temporal Occupancy Grids

Temporal occupancy grids extend the idea of occupancy grids. OGs are a model of a static environment, which is often insufficient to the needs of automated systems which interact with the real world. Instead of keeping track of a single occupancy value for each grid cell, a temporal occupancy grid maintains several such values, each representing the probability of occupancy at a specific time on a specific timescale. For each timescale Δt, the occupancy value at time t is calculated using:

$$O_{i,j,t,\Delta t} = \sum_{t-\Delta t < t' \leq t} R_{i,j,t'} \tag{7}$$

The optimal number of timescales to use in the grid is not something that has yet been studied. We experimented with three timescales in our work to date.

Classification. With the temporal occupancy grid, it is possible to classify the dynamic properties of each cell by considering the occupancy values over different timescales. We can, for example, distinguish between cells containing fixed obstacles (which have high occupancy values on all timescales) and cells which have a high probability of containing moving obstacles (which have high occupancy values on short timescales, but low occupancy values on long timescales).

Let $C_{i,j,t}$ denote the classification for cell (i, j) at time t; we determine $C_{i,j,t}$ based on the occupancy values $O_{i,j,t\Delta t}$ using a simple nearest-neighbor classifier. We define a set of class prototypes $\mathcal{P} = \{\mathcal{P}\}$, in which each prototype P is a set of occupancy values (one for each timescale); for each cell (i, j), we then determine the 'nearest' prototype using a simple Euclidean distance metric:

$$C_{i,j,t} = \arg\min_{P \in \mathcal{P}} \sum_{\Delta t} (O_{i,j,t,\Delta t} - P_{\Delta t}) \tag{8}$$

Consider, for example, a TOG containing three timescales: 1, 15 and 60 seconds. The class prototype for cells containing fixed obstacles would be $P = [1, 1, 1]$, since we expect these cells to be occupied on all timescales. In contrast, the prototype for cells which often contain rapidly moving obstacles would be $P = [1, 0, 0]$.

Collapsing the grid. It is useful to collapse the time dimension of a TOG, generating a map which represents "standard practice" in the area under consideration.

Collapsing the time dimension involves counting the number of times each cell was classified as matching each prototype and then dividing by the number of times the cells were classified. This "mean classification" results in a matrix with two spatial dimensions and an additional number of dimensions equal to the number of timescales being considered.

This map classifies the grid cells in terms of what sort of activity may be expected there in the general case. By comparing this map with individual moments of the TOG, we expect that moments containing unusual activity may be identified.

3.3 Implementation

We implemented a set of post-processing routines for pre-recorded activity data. The data were recorded from several SICK LMS-200 laser rangefinders which had been calibrated so that their relative positions and orientations were known. This calibration was achieved using the mesh relaxation technique described by Howard, Matarić and Sukhatme[4]. Data were stored in a coordinate system which was shared among all lasers.

The first post-processing routine unifies the data from the various viewpoints such that each complete cycle of readings (one reading from each laser) becomes a single image of the world at a particular time. Positive readings ("something is there") override negative readings ("nothing is there"), which in turn override uncertain readings. Stated another way: if any laser sees something, it is assumed that there is indeed an object there, while if some lasers cannot see a location but others can see that it is empty, the location is assumed to be empty. Only if no laser can see a location is it recorded as being uncertainly occupied at that moment. This is reasonable as laser data are not generally prone to false positives.

The second post-processing routine finishes the task of generating the TOG by using a moving window strategy to process the frames which the first post-processor emits. For each timescale under consideration, a window of a different size moves through the data. The value stored in a TOG cell is the average of the occupancies at that location over the period of time equal to the timescale previous to the moment the cell represents.

4 Experimental Validation

The goal of our experiments was to determine to what extent the method could accurately discover the layout and motion profile of a room containing activity, and how well it could localize and classify the activity at any given moment. A final experiment was performed to confirm the real-world utility of the TOG.

4.1 Experimental Design

We performed two experimental scenarios to test our TOG method and a final experiment to assess the utility of the method. In the first experiment, we validated the method's ability to accurately locate features in a static environment. In the second experiment we validated the method's ability to classify the temporal properties of the environment by analyzing scripted human activity. Finally, we assessed the utility of the method by analyzing unscripted human activity in a real-world setting.

For the first experiment, data were gathered from four lasers placed approximately 18 inches above the floor. The environment was entirely static.

For the second experiment, the same configuration of lasers was used. The experiment consisted of a single person moving about the space randomly for approximately 45 seconds followed by approximately three minutes during which four people walked counterclockwise around the border of a four sided figure. The room layout for this experiment is shown in Figure 2. The circular dots mark the locations of the four lasers, while the X-shaped dots mark the corners of the figure that the subjects walked along.

For the third experiment, a single laser was set up in the foyer of a public building and used to record approximately 20 minutes of unscripted activity in the area. The goal of the experiment was to determine whether the often used exterior door would be identifiable in the TOG. Data for the third experiment were gathered by a single laser placed at approximately 18 inches above the floor.

The timescales used during all experiments were 1, 15, and 60 seconds.

4.2 Experimental Results

The first experiment resulted in the correct extraction of the static features of the environment (accurate to within the size of the grid cells). Figure 1 illustrates the result of this experiment.

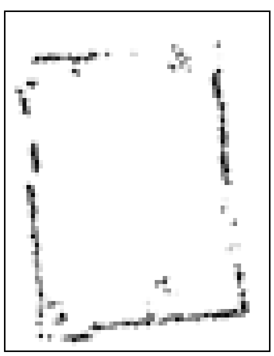


Figure 1: *Probability of cells to be classified as background in experiment 1*

In the second experiment, we were able to extract the time properties of the human activity. Specifically, TOG cells which corresponded to locations on the path that the volunteers walked were classified as having a much higher probability of being occupied on the 1 second timescale than on any other timescale. Figure 3(a) illustrates this result. The figure represents the probability of cells to be classified as occupied primarily on the 1 second timescale. Similar maps of the 15 second and 60 second classifications are essentially empty. Figure 3(d) represents the probability of cells to be classified as background during the same experiment. Note the close correlation with the static features extracted in experiment 1.

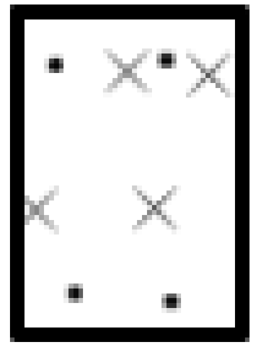


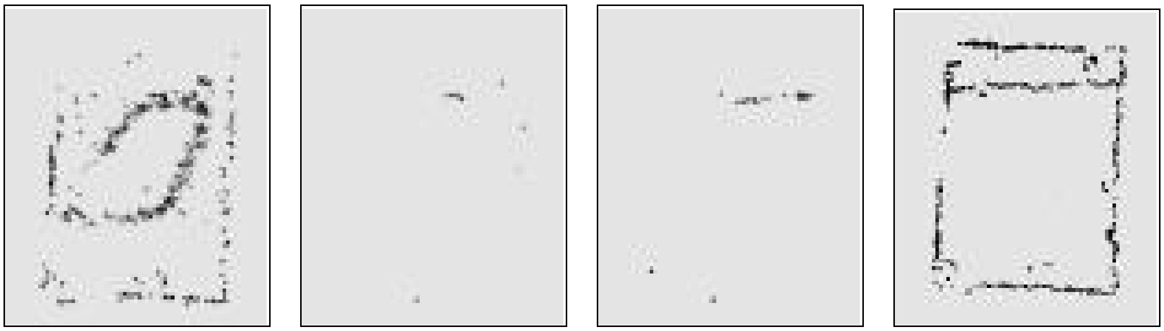
Figure 2: *Layout of the room where the validation experiments occurred*

The final experiment was a test of utility rather than a validation of the method. The goal was to locate the exterior door of the building using the TOG methodology outlined in this paper. The result was that the door was uniquely represented in the collapsed map generated from the TOG. Figures 4(a), 4(b), 4(c) and 4(d) represent the probability of cells to be classified as primarily "occupied on 1 second timescale," "occupied on 15 second timescale," "occupied on 60 second timescale," and "occupied on all timescales" (i.e. background), respectively. The circled feature on Figure 4(c) and Figure 4(d), which is the only feature that appears relatively strongly in both the background and a second timescale, corresponds in position with the exterior door of the building. Figures 4(a) and 4(b) are included to show that the feature does not appear in other timescales.

As a point of interest, Figure 4(a) shows the path between the elevator door, the interior laboratory door, and the elevator as having a relatively high probability of occupancy on the 1 second timescale, while Figure 4(b) shows relatively high probabilities of occupancy on the 15 second timescale at points which correspond at least roughly to the elevator door and to another location where people had a tendency to stop and converse with the experimenter. In order to aid visibility, identical brightness and contrast adjustments have been applied to all four figures representing the foyer experiment.

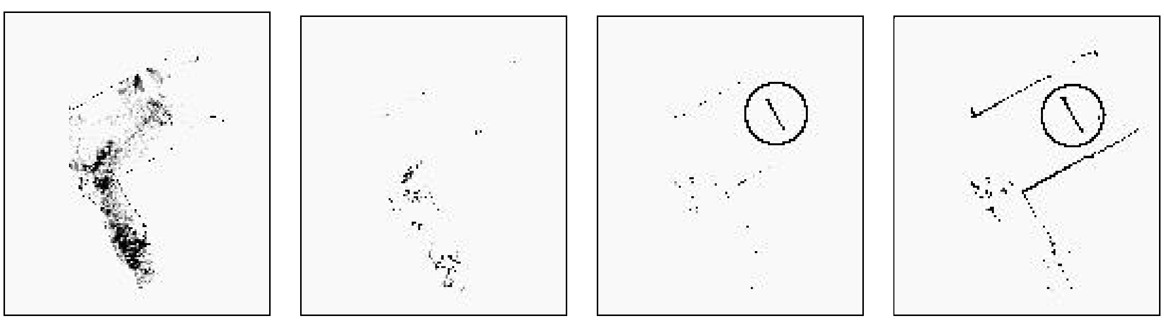
5 Conclusion

This paper introduces the notion of a temporal occupancy grid (TOG) as a tool for classifying motion in an area, both instantaneously and over an extended time. We have demonstrated that TOGs can isolate the location of motion, can classify locations according to the speed of motion that occurs there, and can extract the static features of the environment. This technique has applications in robot navigation, automated security systems, and human activity modeling. In future work we plan to extend and



(a) (b) (c) (d)

__Figure 3:__ Probability of cells to be classified as occupied primarily by (a) 1 second activity, (b) 15 second activity, (c) 60 second activity and (d) "background" during the second experiment



(a) (b) (c) (d)

__Figure 4:__ Probability of cells to be classified as occupied primarily by (a) 1 second activity, (b) 15 second activity, (c) 60 second activity and (d) "background" during the foyer experiment

apply this methodology to the problem of finding unusual and abnormal activity.

Acknowledgments

This work is supported in part by the DARPA MARS Program grant DABT63-99-1-0015 and ONR grant N000140110354.

References

[1] Elfes, A: “Using occupancy grids for mobile robot perception and navigation,” Computer, vol.22, no6, pp.46-57, 1989

[2] Koichi Sato and J.K. Aggarwal: “Tracking and recognizing Two-person interaction in outdoor Image Sequences,” Proceedings of the IEEE Workshop on Multi-Object Tracking, Vancouver, BC, Canada, July 2001

[3] Anurag Mittal and Larry Davis: “Unified Multi-Camera Detection and Tracking Using Region Matching,” Proceedings of the IEEE Workshop on Multi-Object Tracking, Vancouver, BC, Canada, July 2001

[4] Andrew Howard, Maja Matarić and Gaurav Sukhatme: “Relaxation on a mesh: a formalism for generalized localization”, Proceedings,IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1055-1060, Maui, Hawaii, 2001

[5] Dirk Schulz, Wolfram Burgard, Dieter Fox, Armin B. Cremers: “Tracking Multiple Moving Targets with a Mobile Robot using Particle Filters and Statistical Data Association,” Proceedings International Conference on Robotics and Automation (ICRA), pp 1665-1670, Seoul, Korea, 2001

[6] Ting-Hsun Chang and Shaogang Gong: “Tracking Multiple People with a Multi-Camera System,” Proceedings IEEE Workshop on Multi-Object Tracking, Vancouver, BC, Canada, July 2001

[7] Shiloh L. Dockstader and A. Murat Tekalp: “Multiple Camera Fusion for Multi-Object Tracking,” Proceedings IEEE Workshop on Multi-Object Tracking, pp. 95-102, Vancouver, BC, Canada, July 2001

[8] Ronald Mahler: “Engineering Statistics for Multi-Object Tracking,” Proceedings IEEE Workshop on Multi-Object Tracking, Vancouver, BC, Canada, July 2001

[9] Gern, T. and Gilles, E.D: “A multisensor approach to automatic river chart generation,” Intelligent Autonomous Systems (IAS-5), pp 471-478, Amsterdam, Netherlands, 1998

[10] Olson, C.F: “Probabilistic self-localization for mobile robots,” IEEE Transactions on Robotics and Automation, Vol. 16, No. 1, pp 55-66, February 2000.

[11] Robert T. Collins, Alan J. Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt, Lambert Wixson: “A system for Video Surveillance and Monitoring,” tech. report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000

[12] B. Flinchbaugh and T. Bannon: “Autonomous Scene Monitoring System,” Proc. 10th Annual Joint Government-Industry Security Technology Symposium, American Defense Preparedness Association, June 1994